

What is Claimed is:

1. A method of fluorophore bias removal comprising the steps of:

- (a) labeling a first pool of genetic matter, derived from a biological system
- 5 representing a baseline state, with a first fluorophore to obtain a first pool of fluorophore-labeled genetic matter;
- (b) labeling a second pool of genetic matter, derived from a biological system representing a perturbed state, with a second fluorophore to obtain a second pool of fluorophore-labeled genetic matter;
- 10 (c) labeling a third pool of genetic matter, derived from said biological system representing said baseline state, with said second fluorophore to obtain a third pool of fluorophore-labeled genetic matter;
- (d) labeling a fourth pool of genetic matter, derived from said biological system representing said perturbed state, with said first fluorophore to obtain a fourth pool of
- 15 fluorophore-labeled genetic matter;
- (e) contacting said first pool of fluorophore-labeled genetic matter and said second pool of fluorophore-labeled genetic matter with a first microarray under conditions such that hybridization can occur, and determining a first color ratio between said first pool of fluorophore-labeled genetic matter that binds under said conditions to said microarray and
- 20 said second pool of fluorophore-labeled genetic matter that binds under said conditions to said microarray;
- (f) contacting said third pool of fluorophore-labeled genetic matter and said fourth pool of fluorophore-labeled genetic matter with a second microarray under conditions such that hybridization can occur, and determining a second color ratio between said third pool of
- 25 fluorophore-labeled genetic matter that binds under said conditions to said microarray and said fourth pool of fluorophore-labeled genetic matter that binds under said conditions to said microarray; and
- (g) computing an average color ratio by averaging said first color ratio and said second color ratio.

30

2. A computer system for fluorophore bias removal, the computer system comprising a processor, and a memory encoding one or more programs coupled to the processor, wherein the one or more programs cause the processor to perform a method comprising:

- (a) labeling a first pool of genetic matter, derived from a biological system
- 35 representing a baseline state, with a first fluorophore to obtain a first pool of fluorophore-labeled genetic matter;

(b) labeling a second pool of genetic matter, derived from a biological system representing a perturbed state, with a second fluorophore to obtain a second pool of fluorophore-labeled genetic matter;

(c) labeling a third pool of genetic matter, derived from said biological system
5 representing said baseline state, with said second fluorophore to obtain a third pool of fluorophore-labeled genetic matter;

(d) labeling a fourth pool of genetic matter, derived from said biological system representing said perturbed state, with said first fluorophore to obtain a fourth pool of fluorophore-labeled genetic matter;

10 (e) contacting said first pool of fluorophore-labeled genetic matter and said second pool of fluorophore-labeled genetic matter with a first microarray under conditions such that hybridization can occur, and determining a first color ratio between said first pool of fluorophore-labeled genetic matter that binds under said conditions to said microarray and said second pool of fluorophore-labeled genetic matter that binds under said conditions to
15 said microarray;

(f) contacting said third pool of fluorophore-labeled genetic matter and said fourth pool of fluorophore-labeled genetic matter with a second microarray under conditions such that hybridization can occur, and determining a second color ratio between said third pool of fluorophore-labeled genetic matter that binds under said conditions to said microarray and
20 said fourth pool of fluorophore-labeled genetic matter that binds under said conditions to said microarray; and

(g) computing an average color ratio by averaging said first color ratio and said second color ratio.

25 3. A method of fluorophore bias removal, said method comprising determining a color ratio by averaging a first color ratio and a second color ratio wherein said first color ratio and said second color ratio have been determined by:

(a) labeling a first pool of genetic matter, derived from a biological system representing a baseline state, with a first fluorophore to obtain a first pool of fluorophore-
30 labeled genetic matter;

(b) labeling a second pool of genetic matter, derived from a biological system representing a perturbed state, with a second fluorophore to obtain a second pool of fluorophore-labeled genetic matter;

(c) labeling a third pool of genetic matter, derived from said biological system
35 representing said baseline state, with said second fluorophore to obtain a third pool of fluorophore-labeled genetic matter;

fluorophore-labeled genetic matter that binds under said conditions to said microarray and said second pool of fluorophore-labeled genetic matter that binds under said conditions to said microarray; and

- (f) contacting said third pool of fluorophore-labeled genetic matter and said fourth pool of fluorophore-labeled genetic matter with a second microarray under conditions such that hybridization can occur, and determining a second color ratio between said third pool of fluorophore-labeled genetic matter that binds under said conditions to said microarray and said fourth pool of fluorophore-labeled genetic matter that binds under said conditions to said microarray.

10

5. The method of Claim 1 or 3 wherein said first fluorophore and said second fluorophore are selected from the group consisting of Cy2-deoxynucleotide triphosphate, Cy3-deoxynucleotide triphosphate, Cy3.5-deoxynucleotide triphosphate, Cy5-deoxynucleotide triphosphate, Cy5.5-deoxynucleotide triphosphate, Cy7-deoxynucleotide triphosphate, fluorescein, lissamine, phycoerythrin, and rhodamine.

6. The computer system of Claim 2 or 4 wherein said first fluorophore and said second fluorophore are selected from the group consisting of Cy2-deoxynucleotide triphosphate, Cy3-deoxynucleotide triphosphate, Cy3.5-deoxynucleotide triphosphate, Cy5-deoxynucleotide triphosphate, Cy5.5-deoxynucleotide triphosphate, Cy7-deoxynucleotide triphosphate, fluorescein, lissamine, phycoerythrin, and rhodamine.

7. The method of Claim 1 or 3 wherein said first and third pool of genetic matter is cDNA derived by reverse transcription from mRNA extracted from said first biological system.

8. The method of Claim 1 or 3 wherein said second and fourth pool of genetic matter is cDNA derived by reverse transcription from mRNA extracted from said second biological system.

30

9. The method of Claim 1 or 3 wherein said average color ratio is computed by the expression

$$\frac{1}{2} (\log(r_{XY}) - \log(r_{XY}^{(rev)}))$$

- 35 where r_{XY} represents said first color ratio and $r_{XY}^{(rev)}$ represents said second color ratio.

10. The computer system of Claim 2 or 4 wherein said average color ratio is computed by the expression

$$\frac{1}{2} (\log(r_{XY}) - \log(r_{XY}^{(rev)}))$$

5 where r_{XY} represents said first color ratio and $r_{XY}^{(rev)}$ represents said second color ratio.

11. The method of Claim 1 or 3 wherein said average color ratio is plotted against a combined total intensity of said first and second, third, and fourth pool of fluorophore-labeled genetic matter upon hybridization to a third microarray.

12. The computer system of Claim 2 or 4 wherein said average color ratio is plotted against a combined total intensity of said first and second, third, and fourth pool of fluorophore-labeled genetic matter upon hybridization to a third microarray.

13. The method of Claim 1 or 3 wherein said average color ratio is plotted against an intensity metric determined by an amount of intensity generated by fluorophore-labeled genetic matter upon hybridization to a microarray wherein said fluorophore-labeled genetic matter is selected from the group consisting of the first pool of fluorophore-labeled genetic matter, the second pool of fluorophore-labeled genetic matter, the third pool of fluorophore-labeled genetic matter, and the fourth pool of fluorophore-labeled genetic matter.

14. A method for determining a probability that an expression level of a cellular constituent in a plurality of paired differential microarray experiments is altered by a perturbation, wherein each paired differential microarray experiment in said plurality of paired differential microarray experiments comprises a first microarray experiment representing a baseline state of a first biological system, and a second microarray experiment representing a perturbed state of said first biological system, said method comprising the steps of

(a) determining an error distribution statistic by fitting a reference pair of microarray experiments with an intensity independent statistic, wherein said reference pair of microarray experiments comprises a first reference microarray experiment, and a second reference microarray experiment that is a nominal repeat of said first reference microarray experiment;

(b) selecting said cellular constituent from a set of cellular constituents measured in

5 said plurality of paired differential microarray experiments, and, for each paired differential microarray experiment in said plurality of paired differential microarray experiments, determining an amount of change in expression level of said cellular constituent between the second microarray experiment and the first microarray experiment of said paired differential microarray experiment using said error distribution statistic; and

10 (c) determining said probability that said expression level of said cellular constituent in said plurality of paired differential microarray experiments is altered by said perturbation by combining said amount of change in expression level of said cellular constituent determined in step (b) for each paired differential microarray experiment in said plurality of paired differential microarray experiments using a rank based method.

15 15. A computer system for determining a probability that an expression level of a cellular constituent in a plurality of paired differential microarray experiments is altered by a perturbation, wherein each paired differential microarray experiment in said plurality of paired differential microarray experiments comprises a first microarray experiment representing a baseline state of a first biological system, and a second microarray experiment representing a perturbed state of said first biological system; the computer system comprising a processor, and a memory encoding one or more programs coupled to the processor and the one or more programs cause the processor to perform a method

20 comprising the steps of

(a) determining an error distribution statistic by fitting a reference pair of microarray experiments with an intensity independent statistic, wherein said reference pair of microarray experiments comprises a first reference microarray experiment, and a second reference microarray experiment that is a nominal repeat of said first reference microarray

25 experiment;

(b) selecting said cellular constituent from a set of cellular constituents measured in said plurality of paired differential microarray experiments, and, for each paired differential microarray experiment in said plurality of paired differential microarray experiments, determining an amount of change in expression level of said cellular constituent between the

30 second microarray experiment and the first microarray experiment of said paired differential microarray experiment using said error distribution statistic; and

(c) determining said probability that said expression level of said cellular constituent in said plurality of paired differential microarray experiments is altered by said perturbation by combining said amount of change in expression level of said cellular constituent

35 determined in step (b) for each paired differential microarray experiment in said plurality of paired differential microarray experiments using a rank based method.

16. The method of Claim 14 wherein said error distribution statistic is calculated according to a formula

$$\frac{(X - Y)}{\sqrt{\sigma_X^2 + \sigma_Y^2 + f^2(X^2 + Y^2)}}$$

5

where X represents an intensity of a cellular constituent in said first microarray experiment of said reference pair of microarray experiments, Y represents an intensity of said cellular constituent in said second microarray experiment of said reference pair of microarray experiments, σ_X^2 is a variance term for X that represents an additive error level in
10 X, σ_Y^2 is a variance term for Y that represents an additive error level in Y, and f is a fractional multiplicative error level.

17. The computer system of Claim 15 wherein said error distribution statistic is calculated according to a formula

15

$$\frac{(X - Y)}{\sqrt{\sigma_X^2 + \sigma_Y^2 + f^2(X^2 + Y^2)}}$$

where X represents an intensity of a cellular constituent in said first microarray experiment of said reference pair of microarray experiments, Y represents an intensity of
20 said cellular constituent in said second microarray experiment of said reference pair of microarray experiments, σ_X^2 is a variance term for X that represents an additive error level in X, σ_Y^2 is a variance term for Y that represents an additive error level in Y, and f is a fractional multiplicative error level.

25 18. The method of Claim 16 wherein said rank based method comprises determining a rank for said amount of change in expression level of said cellular constituent between said second microarray experiment and said first microarray experiment of said paired differential microarray experiment in relation to all cellular constituents measurements in said plurality of paired differential microarray experiments according to a magnitude derived
30 by the formula of Claim 16.

19. The computer system of Claim 17 wherein said rank based method comprises determining a rank for said amount of change in expression level of said cellular constituent between said second microarray experiment and said first microarray experiment of said
35 paired differential microarray experiment in relation to all cellular constituents measurements in said plurality of paired differential microarray experiments according to a magnitude

derived by the formula of Claim 17.

20. The method of Claim 14 wherein said rank based method determines a probability that a cellular constituent is up-regulated in response to a perturbation.

21. The computer system of Claim 15 wherein said rank based method determines a probability that a cellular constituent is up-regulated in response to a perturbation.

22. The method of Claim 20 wherein said rank based method has the form

$$P(H_0^*) = \prod_i P_i$$

where P_i is said probability that a cellular constituent is up-regulated in said plurality of paired differential microarray experiment i , i is a paired differential microarray experiment selected from said plurality of paired differential microarray experiments, and P is said probability that said expression level of said cellular constituent is up-regulated in response to said perturbation

23. The method of Claim 14 wherein said rank based method determines a probability that a cellular constituent is down-regulated in response to a perturbation.

24. The method of Claim 23 wherein said rank based method has the form

$$P(H_0) = \prod_i (1 - P_i)$$

where P_i is said probability that a cellular constituent is down-regulated in paired differential microarray experiment i , i is selected from said plurality of paired differential microarray experiments, and P is said probability that said cellular constituent is down-regulated in response to said perturbation

25. The method of Claim 14 wherein each paired differential microarray experiment in said plurality of paired differential microarray experiments is a two-fluorophore microarray experiments wherein a first fluorophore represents said baseline state of said biological system and a second fluorophore represents said perturbed state of said biological system.

26. The method of Claim 14 wherein a single fluorophore is used in said paired

experiment pair comprises a first reference microarray experiment and a second reference microarray experiment which is a nominal repeat of said first reference microarray experiment;

- (b) determining an amount of differential expression of said cellular constituent a plurality of times;
- (c) for each amount of differential expression determined in accordance with (b), calculating a corresponding amount of error based on a magnitude derived by said error distribution statistic; and
- (d) computing said weighted mean differential intensity by inversely weighting each said amount of differential expression of said cellular constituent determined in step (b) by the corresponding amount of error determined in step (c) according to the formula

$$x = \frac{\sum (x_i / \sigma_i^2)}{\sum (1 / \sigma_i^2)}$$

- where x is said weighted mean differential intensity of in said expression level of said cellular constituent, x_i is an amount of differential expression of said cellular constituent i and σ_i^2 is a corresponding error for x_i .

30. The method of Claim 28, wherein step (b) further comprises:

- (i) measuring a first intensity of a position on a microarray after said microarray has been contacted with a first pool of fluorophore-labeled genetic matter derived from a biological system that represents a baseline state, wherein said position on said microarray represents said cellular constituent;
- (ii) measuring a second intensity of said position on a microarray after said microarray has been incubated with a second pool of fluorophore-labeled genetic matter derived from a biological system that represents a perturbed state; and
- (iii) computing said differential expression of said cellular constituent by subtracting said second intensity from said first intensity.

31. The method of Claim 28, wherein said first pool and said second pool of fluorophore-labeled genetic matter comprises cDNA derived from mRNA by reverse transcription.

32. The method of Claim 28 wherein said error distribution statistic is calculated according to the formula

$$\frac{(X - Y)}{\sqrt{\sigma_X^2 + \sigma_Y^2 + f^2(X^2 + Y^2)}}$$

- where X represents an intensity of a cellular constituent in said first microarray experiment of said reference microarray experiment pair, Y represents an intensity of said cellular constituent in said second microarray experiment of said reference microarray experiment pair, α_x^2 is a variance term for X that represents an additive error level in X, α_y^2 is a variance term for Y that represents an additive error level in Y, and f is a fractional multiplicative error level.

33. A method for determining a confidence of a weighted average of a plurality of cellular constituent differential expression measurements determined for a predetermined cellular constituent, j, wherein each cellular constituent differential expression measurement is determined by a paired differential microarray experiment selected from a plurality of paired differential microarray experiments wherein each paired differential microarray experiment comprises a first microarray experiment representing a baseline state of a biological system and a second microarray experiment representing a perturbed state of a biological system, said method comprising the steps of

- (a) determining an error distribution statistic by fitting a reference pair of microarray experiments with an intensity independent statistic, wherein said reference pair of microarray experiments comprises a first reference microarray experiment and a second reference microarray experiment that is a nominal repeat of said first reference microarray experiment;

- (b) for each paired differential microarray experiment in said plurality of paired differential microarray experiments, determining an amount of error based upon said error distribution statistic;

- (c) determining a scatter s_j for cellular constituent j based upon the plurality of paired differential microarray experiments using a relationship

$$s_j \cong \frac{1}{N-1} \sum_i (x_i - \bar{x})^2$$

- where x_i is a differential measurement of cellular constituent j that is determined by paired differential microarray experiment i, \bar{x} is the unweighted mean value of all differential measurements of cellular constituent j in said plurality of paired differential microarray experiments, and N is a number of paired differential microarray experiments in said plurality of paired differential microarray experiments; and

- (d) combining said amount of error for each paired differential microarray experiment determined in step (b) with said scatter s_j to determine said confidence of said

weighted average of said plurality of cellular constituent differential expression measurements determined for said predetermined cellular constituent j.

34. The method of Claim 33 wherein said error distribution statistic is calculated
5 according to the formula

$$\frac{(X - Y)}{\sqrt{\sigma_X^2 + \sigma_Y^2 + f^2(X^2 + Y^2)}}$$

- where X represents an intensity of a cellular constituent in said first microarray
10 experiment of said reference pair of microarray experiments, Y represents an intensity of said cellular constituent in said second microarray experiment of said reference pair of microarray experiments, σ_X^2 is a variance term for X that represents an additive error level in X, σ_Y^2 is a variance term for Y that represents an additive error level in Y, and f is a fractional multiplicative error level.

15

35. The method of Claim 33 wherein each cellular constituent differential expression measurement in said plurality of cellular constituent differential expression measurements for cellular constituent j is determined by

- (i) measuring a first intensity of a position on a microarray after said microarray has
20 been contacted with a first pool of fluorophore-labeled genetic matter derived from a biological system that represents a baseline state wherein said position on said microarray corresponds to said cellular constituent j;
(ii) measuring a second intensity of a position on a microarray after said microarray has been incubated with a second pool of fluorophore-labeled genetic matter derived from a
25 biological system that represents a perturbed state wherein said position on said microarray corresponds to said cellular constituent j; and
(iii) computing said cellular constituent differential expression by subtracting said second intensity from said first intensity.

30

36. The method of Claim 33, wherein step (b) further comprises:

- (i) plotting said error statistic on an X-Y graph wherein a first axis represents intensity and a second axis represent an expression ratio; and
(ii) determining said amount of error by identifying a position along said first axis by plotting said second intensity on said first axis and measuring a width based on $\pm 1\sigma$ grid
35 lines plotted according to said error statistic at said position.

37. The method of Claim 33, wherein step (d) further comprises combining said amount of error for each paired differential microarray experiment determined in step (b) of Claim 33 with said scatter s_j of step (c) of Claim 33 according to a formula:

$$\sigma_x = \frac{1}{N} \left[\left(\sqrt{\frac{1}{\sum_i \frac{1}{\sigma_i^2}}} \right) + (N-1) * s_j \right]$$

where

$$\frac{1}{\sum (\frac{1}{\sigma_i^2})}$$

is determined by said error distribution statistic in accordance with step (b) of Claim 33, N is a number of paired differential microarray experiments used to calculate s_j and σ_x is a representation of said confidence of said weighted average of said plurality of cellular constituent differential expression measurements determined for said predetermined cellular constituent j.

38. A computer system for determining a confidence of a weighted average of a plurality of cellular constituent differential expression measurements determined for a predetermined cellular constituent, j, wherein each cellular constituent differential expression measurement is determined by a paired differential microarray experiment selected from a plurality of paired differential microarray experiments wherein each paired differential microarray experiment comprises a first microarray experiment representing a baseline state of a biological system and a second microarray experiment representing a perturbed state of a biological system; wherein the computer system comprising a processor, and a memory encoding one or more programs coupled to the processor and the one or more programs cause the processor to perform a method comprising the steps of

(a) determining an error distribution statistic by fitting a reference pair of microarray experiments with an intensity independent statistic, wherein said reference pair of microarray experiments comprises a first reference microarray experiment and a second reference microarray experiment that is a nominal repeat of said first reference microarray experiment;

(b) for each paired differential microarray experiment in said plurality of paired differential microarray experiments, determining an amount of error based upon said error distribution statistic;

(c) determining a scatter s_j for cellular constituent j based upon the plurality of paired

differential microarray experiments using a relationship

$$s_j \equiv \frac{1}{N-1} \sum_i (x_i - \bar{x})^2$$

5 where x_i is a differential measurement of cellular constituent j that is determined by paired differential microarray experiment i , \bar{x} is the unweighted mean value of all differential measurements of cellular constituent j in said plurality of paired differential microarray experiments, and N is a number of paired differential microarray experiments in said plurality of paired differential microarray experiments; and

10 (d) combining said amount of error for each paired differential microarray experiment determined in step (b) with said scatter s_j to determine said confidence of said weighted average of said plurality of cellular constituent differential expression measurements determined for said predetermined cellular constituent j .

15 39. The computer system of Claim 38 wherein said error distribution statistic is calculated according to the formula

$$\frac{(X - Y)}{\sqrt{\sigma_X^2 + \sigma_Y^2 + f^2(X^2 + Y^2)}}$$

20 where X represents an intensity of a cellular constituent in said first microarray experiment of said reference pair of microarray experiments, Y represents an intensity of said cellular constituent in said second microarray experiment of said reference pair of microarray experiments, σ_X^2 is a variance term for X that represents an additive error level in X , σ_Y^2 is a variance term for Y that represents an additive error level in Y , and f is a
25 fractional multiplicative error level.

40. The computer system of Claim 38 wherein each cellular constituent differential expression measurement in said plurality of cellular constituent differential expression measurements for cellular constituent j is determined by

30 (i) measuring a first intensity of a position on a microarray after said microarray has been contacted with a first pool of fluorophore-labeled genetic matter derived from a biological system that represents a baseline state wherein said position on said microarray corresponds to said cellular constituent j ;

(ii) measuring a second intensity of a position on a microarray after said microarray
35 has been incubated with a second pool of fluorophore-labeled genetic matter derived from a biological system that represents a perturbed state wherein said position on said microarray

corresponds to said cellular constituent j; and

(iii) computing said cellular constituent differential expression by subtracting said second intensity from said first intensity.

5 41. The computer system of Claim 38, wherein step (b) further comprises:

(i) plotting said error statistic on an X-Y graph wherein a first axis represents intensity and a second axis represent an expression ratio; and

(ii) determining said amount of error by identifying a position along said first axis by plotting said second intensity on said first axis and measuring a width based on $\pm 1\sigma$ grid
10 lines plotted according to said error statistic at said position.

42. The computer system of Claim 38, wherein step (d) further comprises combining said amount of error for each plurality of paired differential microarray experiments determined in step (b) of Claim 38 with said scatter s_j of step (c) of Claim 38 according to a
15 formula:

$$\sigma_x = \frac{1}{N} \left[\sqrt{\frac{1}{\sum \frac{1}{\sigma_i^2}}} + (N-1) * s_j \right]$$

20 where

$$\frac{1}{\sum \left(\frac{1}{\sigma_i^2} \right)}$$

is determined by said error distribution statistic in accordance with step (b) of Claim
25 38, N is a number of paired differential microarray experiments used to calculate s_j and σ_x is a representation of said confidence of said weighted average of said plurality of cellular constituent differential expression measurements determined for said predetermined cellular constituent j.

30

add
A10

35